

Equivariant adjusted least squares estimator in two-line fitting model

Sergiy Shklyar

Taras Shevchenko National University of Kyiv, Ukraine

shklyar@univ.kiev.ua (S. Shklyar)

Received: 30 January 2016, Revised: 19 February 2016, Accepted: 19 February 2016,
 Published online: 21 March 2016

Abstract We consider the two-line fitting problem. True points lie on two straight lines and are observed with Gaussian perturbations. For each observed point, it is not known on which line the corresponding true point lies. The parameters of the lines are estimated.

This model is a restriction of the conic section fitting model because a couple of two lines is a degenerate conic section. The following estimators are constructed: two projections of the adjusted least squares estimator in the conic section fitting model, orthogonal regression estimator, parametric maximum likelihood estimator in the Gaussian model, and regular best asymptotically normal moment estimator.

The conditions for the consistency and asymptotic normality of the projections of the adjusted least squares estimator are provided. All the estimators constructed in the paper are equivariant. The estimators are compared numerically.

Keywords Conic section fitting, curve fitting, subspace clustering

2010 MSC 62J05, 62H12, 62H30

1 Introduction

1.1 Two-line fitting model

Consider a problem of estimation of two lines by perturbed observations of points that lie on the lines. Let the true points (ξ_i, η_i) lie on the union of two different lines $\eta = k_1\xi + h_1$ and $\eta = k_2\xi + h_2$, that is,

$$\begin{cases} \text{either} & \eta_i = k_1\xi_i + h_1, \\ \text{or} & \eta_i = k_2\xi_i + h_2, \end{cases} \quad i = 1, 2, \dots \quad (1)$$

© 2016 The Author(s). Published by VTeX. Open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.

Let these points be observed with perturbations $(\delta_i, \varepsilon_i)$, $i = 1, \dots, n$, that is, the observed points are (x_i, y_i) , $i = 1, \dots, n$, with

$$x_i = \xi_i + \delta_i, \quad (2)$$

$$y_i = \eta_i + \varepsilon_i. \quad (3)$$

The perturbations are assumed to be independent and identically normally distributed,

$$\begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix} \sim N(0, \sigma^2 \mathbf{I}), \quad (4)$$

where \mathbf{I} is the 2×2 identity matrix.

The parameters k_1, h_1, k_2, h_2 , and σ^2 are to be estimated.

We consider both functional and structural models. In *functional* model, the true points are assumed to be nonrandom. In *structural* model, the true points are assumed to be independent and identically distributed (i.i.d.). The errors $(\delta_i, \varepsilon_i)$ are i.i.d. and independent of the true points.

In the structural model, $(\xi_i, \eta_i, \delta_i, \varepsilon_i)$ are i.i.d. random vectors, and thus, the observed points (x_i, y_i) are i.i.d. In the functional model, the observed points are independent, Gaussian, with different means but with common covariance matrix.

Remark 1. The true lines defined by Eqs. (1) cannot be parallel to the y -axis. In order to avoid overflows during evaluation of the estimators (except of RBAN-moment estimator), another parameterization is used internally: $\vec{\tau}^\top (\zeta - \zeta_0) = 0$, where $\vec{\tau}$ is a unit vector orthogonal to the line, and ζ_0 is a point on the line. The computation of the RBAN-moment estimator (see Section 2.4) is implemented for explicit parameterization only. Computational optimization of the RBAN-moment estimator is a matter of further work.

The explicit parameterization has the advantage that the number of parameters is equal to the dimension of parameter space. (In [12], the second-order equation (5) has six unknown coefficients, but the conic section can be parameterized with five parameters. The parameter space for the parameters of the conic section was the five-dimensional unit sphere in the six-dimensional Euclidean space. Mismatch between the number of parameters and the dimension of the parameter space made the asymptotic covariance matrix of the estimator *singular*.)

In simulations, the confidence intervals for the coordinates of the intersection point of the two lines are obtained based on the asymptotic covariance matrix for the intersection point. For the projections the ALS2 estimator, that asymptotic covariance matrix can be evaluated without use of explicit line parameterization.

1.2 Conic section fitting model

Let the true points (ξ_i, η_i) lie on the second-order algebraic curve

$$A\xi_i^2 + 2B\xi_i\eta_i + C\eta_i^2 + 2D\xi_i + 2E\eta_i + F = 0, \quad i = 1, 2, \dots \quad (5)$$

Hereafter, a second-order algebraic curve is called a “conic section” or a “conic.”

The points are observed with Gaussian perturbations, and the perturbed points are denoted as (x_i, y_i) . We have the same equations

$$\begin{aligned} x_i &= \xi_i + \delta_i, \\ y_i &= \eta_i + \varepsilon_i, \end{aligned} \quad \begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix} \sim N(0, \sigma^2 \mathbf{I}),$$

as (2)–(4) in the two-line fitting model.

The vector of coefficients in (5) is denoted by $\beta = (A, 2B, C, 2D, 2E, F)^\top$. The nonzero vector β and the error variance σ^2 are the parameters of interest.

Similarly to the two-line fitting model, the *functional* and the *structural* models are distinguished.

A couple of lines is a degenerate case of a conic section. Therefore, the conic section fitting model is an extension of the two-line fitting model.

1.3 ALS2 estimator in conic section fitting model

We consider the adjusted least squares (ALS) estimator for unknown σ^2 . The estimator is constructed in [7]. Introduce the 6×6 symmetric matrix

$$\begin{aligned} &\psi(x, y; v) \\ &= \begin{pmatrix} x^4 - 6x^2v + 3v^2 & (x^3 - 3xv)y & (x^2 - v)(y^2 - v) & * & * & * \\ (x^3 - 3xv)y & (x^2 - v)(y^2 - v) & x(y^3 - 3yv) & * & * & * \\ (x^2 - v)(y^2 - v) & x(y^3 - 3yv) & y^4 - 6y^2v + 3v^2 & * & * & * \\ x^3 - 3xv & (x^2 - v)y & x(y^2 - v) & x^2 - v & xy & x \\ (x^2 - v)y & x(y^2 - v) & y^3 - 3yv & xy & y^2 - v & y \\ x^2 - v & xy & y^2 - v & x & y & 1 \end{pmatrix}. \end{aligned}$$

Asterisks are typed instead of some entries above the diagonal of a symmetric matrix. The entries of the matrix $\psi(x, y; v)$ are generalized Hermite polynomials in x and y . The matrix $\psi(x, y; v)$ is constructed such that $E\psi(x_i, y_i; \sigma^2) = \psi(\xi_i, \eta_i; 0)$ in the functional model and $\psi(\xi_i, \eta_i; 0)\beta = 0$ for the true points and true parameters.

Denote

$$\Psi_n(v) = \sum_{i=1}^n \psi(x_i, y_i; v).$$

The estimator $\hat{\sigma}^2$ of the error variance σ^2 is obtained from the equation

$$\lambda_{\min}(\Psi_n(\hat{\sigma}^2)) = 0. \quad (6)$$

Equation (6) always has a unique nonnegative solution. If $n \geq 6$, then the solution to (6) is positive almost surely.

The matrix $\Psi_n(\hat{\sigma}^2)$ is singular. Define the estimator $\hat{\beta}$ of the vector β as a nonzero solution to the equation

$$\Psi_n(\hat{\sigma}^2)\hat{\beta} = 0.$$

The strong consistency of the ALS2 estimator is proved in [7] and [11] under somewhat different conditions. The asymptotic normality is proved in [12] for the functional model and in [13] for the structural model. Two consistent estimators of the asymptotic covariance matrix are constructed in [13].

Denote

$$\begin{aligned}
 \psi'_v(x, y; v) &= \frac{\partial}{\partial v} \psi(x, y; v), & \Psi''_1 &= \frac{\partial^2}{\partial v^2} \psi(x, y; v), \\
 \bar{\Psi}'_n(v) &= \frac{d}{dv} \Psi_n(v), \\
 \bar{\Psi}_n &= \sum_{i=1}^n \psi(\xi_i, \eta_i; 0), & \bar{\Psi}'_n &= \sum_{i=1}^n \psi'_v(\xi_i, \eta_i; 0), \\
 \bar{\Psi}_\infty &= \lim_{n \rightarrow \infty} \frac{1}{n} \bar{\Psi}_n = \lim_{n \rightarrow \infty} \frac{1}{n} \Psi_n(\sigma^2), & \bar{\Psi}'_\infty &= \lim_{n \rightarrow \infty} \frac{1}{n} \bar{\Psi}'_n = \lim_{n \rightarrow \infty} \frac{1}{n} \Psi'_n(\sigma^2).
 \end{aligned}$$

Under the conditions of Proposition 1 stated further, the latter limits exist almost surely. See [12] for explicit expressions of the matrices $\psi'_v(x, y; v)$, Ψ''_1 , $\bar{\Psi}_\infty$, and $\bar{\Psi}'_\infty$. Note that Ψ''_1 is a constant matrix.

Proposition 1. *In the functional model, for all integer $p \geq 0$ and $q \geq 0$ such that $p + q \leq 4$, let the following limits exist and be finite:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi_i^p \eta_j^q =: \mu_{p,q},$$

whereas in the structural model, let $E \xi_1^4 < \infty$ and $E \eta_1^4 < \infty$. In both models, let $\text{rank } \bar{\Psi}_\infty = 5$. Then:

1. The estimator $\hat{\beta}$ is strongly consistent in the following sense:

$$\min \left(\left\| \frac{\hat{\beta}}{\|\hat{\beta}\|} - \frac{\beta}{\|\beta\|} \right\|, \left\| \frac{\hat{\beta}}{\|\hat{\beta}\|} + \frac{\beta}{\|\beta\|} \right\| \right) \rightarrow 0 \quad \text{a.s.}, \quad (7)$$

$$\hat{\sigma}^2 \rightarrow \sigma^2 \quad \text{a.s.} \quad (8)$$

2. $\beta^\top \bar{\Psi}'_\infty \beta < 0$.

3. Eventually, $\hat{\beta}^\top \Psi'_n(\hat{\sigma}^2) \hat{\beta} < 0$.

“Eventually” in the previous statement means that almost surely there exists n_0 such that $\hat{\beta}^\top \Psi'_n(\hat{\sigma}^2) \hat{\beta} < 0$ for all $n \geq n_0$. In other words, almost surely, $\hat{\beta}^\top \Psi'_n(\hat{\sigma}^2) \hat{\beta} \geq 0$ holds only for finitely many n .

Denote the normalized version of the true parameter

$$\beta_{\text{tn}} = \sqrt{\frac{-1}{\beta^\top \bar{\Psi}'_\infty \beta}} \beta.$$

Normalize the estimator of β in such a way that $\tilde{\beta}^\top \Psi'_n(\hat{\sigma}^2) \tilde{\beta} = -n$ and $\beta^\top \tilde{\beta} \geq 0$. Therefore, denote

$$\tilde{\beta} = \begin{cases} \sqrt{\frac{-n}{\beta^\top \Psi'_n(\hat{\sigma}^2) \beta}} \hat{\beta} & \text{if } \beta^\top \hat{\beta} \geq 0, \\ -\sqrt{\frac{-n}{\beta^\top \Psi'_n(\hat{\sigma}^2) \beta}} \hat{\beta} & \text{if } \beta^\top \hat{\beta} < 0. \end{cases} \quad (9)$$

Proposition 2. 1. Under the conditions of Proposition 1, the estimator $\tilde{\beta}$ is a strongly consistent estimator of $\beta_{\text{tn}} = (-\beta^\top \bar{\Psi}'_\infty \beta)^{-1/2} \beta$, that is, $\tilde{\beta} \rightarrow \beta_{\text{tn}}$ a.s.

2. In the functional model, for all integer $p \geq 0$ and $q \geq 0$ such that $p + q \leq 6$, let the following limits exist and be finite:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi_i^p \eta_j^q =: \mu_{p,q},$$

whereas in the structural model, let $E \xi_1^6 < \infty$ and $E \eta_1^6 < \infty$. In both models, let $\text{rank } \bar{\Psi}_\infty = 5$. Then the estimator $\hat{\theta} = (\tilde{\beta}^\top, \hat{\sigma}^2)^\top$ is asymptotically normal in the following sense:

$$\sqrt{n} \begin{pmatrix} \tilde{\beta} - \beta_{\text{tn}} \\ \hat{\sigma}^2 - \sigma^2 \end{pmatrix} \xrightarrow{d} N(0, \Sigma_{\hat{\theta}}), \quad (10)$$

where

$$\begin{aligned} \Sigma_{\hat{\theta}} &= \begin{pmatrix} \bar{\Psi}_\infty & \bar{\Psi}'_\infty \beta_{\text{tn}} \\ \beta_{\text{tn}}^\top \bar{\Psi}'_\infty & \frac{1}{2} \beta_{\text{tn}}^\top \bar{\Psi}''_\infty \beta_{\text{tn}} \end{pmatrix}^{-1} \mathbf{B} \begin{pmatrix} \bar{\Psi}_\infty & \bar{\Psi}'_\infty \beta_{\text{tn}} \\ \beta_{\text{tn}}^\top \bar{\Psi}'_\infty & \frac{1}{2} \beta_{\text{tn}}^\top \bar{\Psi}''_\infty \beta_{\text{tn}} \end{pmatrix}^{-1}, \\ \mathbf{B} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E s_i s_i^\top, \\ s_i &= \begin{pmatrix} \psi(x_i, y_i; \sigma^2) \beta_{\text{tn}} \\ \frac{1}{2} \beta_{\text{tn}}^\top \psi(x_i, y_i; \sigma^2) \beta_{\text{tn}} - \frac{1}{2} \end{pmatrix}. \end{aligned}$$

3. Under the conditions of part 2 of Proposition 2, the following estimator of the asymptotic covariance matrix is consistent:

$$\begin{aligned} \hat{\Sigma}_{\hat{\theta}}(n) &= \mathbf{A}^{-1}(n) \mathbf{B}(n) \mathbf{A}^{-1}(n), \\ \mathbf{A}(n) &= \begin{pmatrix} \frac{1}{n} \bar{\Psi}_n(\hat{\sigma}^2) & \frac{1}{n} \bar{\Psi}'_n(\hat{\sigma}^2) \tilde{\beta} \\ \frac{1}{n} \tilde{\beta}^\top \bar{\Psi}'_n(\hat{\sigma}^2) & \frac{1}{2} \tilde{\beta}^\top \bar{\Psi}''_n(\hat{\sigma}^2) \tilde{\beta} \end{pmatrix}, \quad \mathbf{B}(n) = \frac{1}{n} \sum_{i=1}^n \hat{s}_i \hat{s}_i^\top, \\ \hat{s}_i &= \begin{pmatrix} \psi(x_i, y_i; \hat{\sigma}^2) \tilde{\beta} \\ \frac{1}{2} \tilde{\beta}^\top \psi(x_i, y_i; \hat{\sigma}^2) \tilde{\beta} - \frac{1}{2} \end{pmatrix}, \end{aligned}$$

that is,

$$\hat{\Sigma}_{\hat{\theta}}(n) \rightarrow \Sigma_{\hat{\theta}}$$

in probability.

1.4 Estimation methods

The methods of fitting an algebraic curve (or surface) to observed points can be classified as follows.

Algebraic distance methods, where the residuals in the equations for the algebraic curve are minimized. For example, the minimum point of the sum of squared residuals $\sum_{i=1}^n (Ax_i^2 + 2Bx_iy_i + Cy_i^2 + 2Dx_i + 2Ey_i + F)^2$ (with some normalizing constraint in order to avoid $A = B = \dots = F = 0$) in the conic fitting problem and $\sum_{i=1}^n (k_1x_i + h_1 - y_i)^2 (k_2x_i + h_2 - y_i)^2$ in the two-line fitting problem is called the *ordinary least squares* (OLS) estimator.

The criterion function for the OLS estimator is simple enough and can be adjusted so that the resulting estimator is consistent (under some conditions). Such an estimator is called the *adjusted least squares* (ALS) estimator. The OLS and ALS estimators are method-of-moments estimators, meaning that the criterion functions for the estimators are polynomials whose coefficients are sample moments of coordinates of the observed points. Hence, the OLS and ALS estimators can be computed efficiently.

In order to obtain parameters of two lines, the observed points are fitted with a conic section, and then the parameters of the conic section are used to obtain the parameters of two lines. There are some papers where this idea is used.

The problem of estimating the fundamental matrix for two-camera view is considered in [6]. The fundamental matrix is a singular matrix whose left and right null-vectors are the coordinates of each camera in the coordinate system of the other camera. Initially, the ALS estimator of the fundamental matrix is evaluated. Then it is projected so that the estimated fundamental matrix is singular.

In [14] the problem of segmentation of a finite-dimensional vector space onto linear subspaces is considered, and the generalized principal component analysis method is introduced. The sample is fitted with an algebraic cone (a set of points that satisfy a homogeneous algebraic equation) by the OLS method. Then subspaces are extracted from the algebraic cone with use of a small learning sample. An application of segmentation of a vector space onto hyperplanes for searching planes on binocular image is given in [16].

In [15] an ellipsoid fitting problem with a constraint such that a center of the ellipsoid lies on a given line is considered. The algebraic distance with embedded constraint is minimized. The analytical (behavioral) properties of the optimization problem are studied. We consider a conic section fitting problem but with different constraint—the conic is degenerated to a couple of straight lines.

Geometric distance methods, where distances between the estimated curve and each point are minimized. The sum of squares of those distances is minimized, and the *orthogonal regression* (OR) estimator is obtained.

A numerical algorithm for evaluation of the orthogonal regression estimator is presented in monograph [1].

The orthogonal regression is consistent in the single straight line fitting problem [3, Section 1.3.2(a)]. In nonlinear models, the estimator may be inconsistent. There is a one-step correction procedure in explicit and implicit models [5, 9] with application in the ellipsoid fitting model [9]. However, in the two-line fitting model, the correction from [9] is unstable.

Probabilistic methods. They are used to obtain the *maximum likelihood* (ML) estimator and Bayes estimators.

1.5 Notation

Let $\{A_n, n = 1, 2, \dots\}$ be a sequence of random events. The random event A_n is said to hold eventually if almost surely there exists n_0 such that A_n occurs for all $n \geq n_0$. In other words, the random event A_n holds eventually if and only if it does not occur only for finitely many n almost surely.

The estimator $\hat{\beta}$ is called asymptotically normal if $\sqrt{n}(\hat{\beta} - \beta_{\text{true}}) \rightarrow N(0, \Sigma)$ in distribution, where the asymptotic covariance matrix Σ may be singular, and n is the sample size. This definition differs from the conventional one adopted in asymptotic theory because here only \sqrt{n} -asymptotic normality is considered.

Let $\zeta \sim N(\mu, \Sigma)$ be a bivariate random vector. Then $E = \{z : (z - \mu)^\top \Sigma^{-1}(z - \mu) \leq 1\}$ is called the 40% ellipsoid of the normal distribution because $P(\zeta \in E) \approx 0.3935$. This is the ellipsoid where the probability density function is at least 0.3679 of its maximum.

1.6 Outline

In Section 2, we construct five estimators for parameters of the two line fitting model. In Section 3, we propose two definitions of the equivariance of an estimator and state that all of the five estimators are equivariant. The estimators are compared numerically in Section 4. The proofs are given in Appendix A.

2 Estimators

2.1 ALS2 estimator and its projections

The two-line fitting model is a restriction of the conic section fitting model. A couple of lines defined by the equation $(k_1\xi - \eta + h_1)(k_2\xi - \eta + h_2) = 0$ is a degenerate conic section

$$A\xi^2 + 2B\xi\eta + C\eta^2 + 2D\xi + 2E\eta + F = 0, \quad (11)$$

with coefficients

$$\begin{aligned} A &= Ck_1k_2, & 2D &= C(k_1h_2 + k_2h_1), \\ 2B &= -C(k_1 + k_2), & 2E &= -C(h_1 + h_2), \\ F &= Ch_1h_2, \end{aligned} \quad (12)$$

with a constraint $C \neq 0$.

The conic section ALS2 estimator provides estimation of the error variance σ^2 and the coefficients A, B, \dots, F .

Denote by $\nu(i) \in \{1, 2\}$ the indicator of a line which the true point (ξ_i, η_i) belongs to. Equation (1) can be rewritten as

$$\eta_i = k_{\nu(i)}\xi_i + h_{\nu(i)}, \quad i = 1, 2, \dots$$

The indicator $\nu(i)$ is nonrandom in the functional model, and it is a random variable in the structural model.

Proposition 3. *Let, in the functional model,*

$$\sum_{n=1}^{\infty} \frac{\xi_n^6}{n^2} < \infty;$$

$$\text{either } k_1 \neq k_2 \text{ or } \begin{cases} h_1 \neq h_2, \\ \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n \xi_i^2 < \infty; \text{ and} \end{cases}$$

$$\liminf_{n \rightarrow \infty} \lambda_{\min} \left(\frac{1}{n} \sum_{\substack{i=1, \dots, n \\ \nu(i)=j}} \begin{pmatrix} 1 & \xi_i & \xi_i^2 \\ \xi_i & \xi_i^2 & \xi_i^3 \\ \xi_i^2 & \xi_i^3 & \xi_i^4 \end{pmatrix} \right) > 0 \text{ for } j = 1, 2.$$

Then the ASL2 estimators $\hat{\beta}$ and $\hat{\sigma}^2$ are strongly consistent in the sense of (7) and (8).

There are two cases where the structural model is not identifiable. If the common distribution of the true points is concentrated on a straight line and on a single point (presumably not on the line), that is,

$$\exists \text{ line } \ell \subset \mathbb{R}^2 \exists z \in \mathbb{R}^2 : \text{supp}(\xi_1, \eta_1) \subset \ell \cup \{z\}, \quad (13)$$

then there are many ways to fit the true points with two lines. If the common distribution of the true points is concentrated in four points, that is,

$$\# \text{supp}(\xi_1, \eta_1) = 4, \quad (14)$$

then there are three ways to fit the true points with two lines (unless three of the four points lie on a straight line, which is a particular case of (13)).

Proposition 4. *In the structural model, assume that $E|\xi_1|^3 < \infty$ and that nonidentifiability conditions (13) and (14) do not hold. Then the ALS2 estimator is strongly consistent in the sense of (7) and (8).*

In order to estimate the parameters k_1 , h_1 , k_2 , and h_2 , we can solve Eqs. (12). With ignoring the last equation $F = Ch_1h_2$, the solution is

$$k_{1,2} = \frac{-B \pm \sqrt{B^2 - AC}}{C}, \quad (15)$$

$$h_1 = \frac{2(D + k_1E)}{C(k_2 - k_1)}, \quad (16)$$

$$h_2 = \frac{2(D + k_2E)}{C(k_1 - k_2)}. \quad (17)$$

Substituting the elements of the ALS2 estimator $\hat{\beta} = (\hat{A}, 2\hat{B}, \hat{C}, 2\hat{D}, 2\hat{E}, \hat{F})^\top$ into the right-hand side of (15)–(17), we obtain an “ignore- \hat{F} ” estimator:

$$\hat{k}_{1,2} = \frac{-\hat{B} \pm \sqrt{\hat{B}^2 - \hat{A}\hat{C}}}{\hat{C}}, \quad (18)$$

$$\hat{h}_1 = \frac{2(\hat{D} + \hat{k}_1\hat{E})}{\hat{C}(\hat{k}_2 - \hat{k}_1)}, \quad (19)$$

$$\hat{h}_2 = \frac{2(\hat{D} + \hat{k}_2 \hat{E})}{\hat{C}(\hat{k}_1 - \hat{k}_2)}. \quad (20)$$

If the conic section estimated by the ALS2 estimator is a hyperbola, then the “ignore- \hat{F} ” estimate of the two lines comprises the asymptotes of the hyperbola.

Choose the sign \pm in (18) such that $\hat{k}_1 < \hat{k}_2$.

We need the notation

$$(k_1, h_1, k_2, h_2)^\top = \text{lob}(\beta)$$

for the function that expresses the line parameters k_1, h_1, k_2, h_2 in elements of β and is defined by (15)–(17). With this notation, we can write

$$(\hat{k}_1, \hat{h}_1, \hat{k}_2, \hat{h}_2)^\top = \text{lob}(\hat{\beta}) = \text{lob}(\tilde{\beta}).$$

Proposition 5. *In the functional model, assume the following:*

$$\begin{aligned} & k_1 < k_2, \\ & \sum_{n=1}^{\infty} \frac{\xi_n^6}{n^2} < \infty, \quad \text{and} \\ & \liminf_{n \rightarrow \infty} \lambda_{\min} \left(\frac{1}{n} \sum_{\substack{i=1, \dots, n \\ \nu(i)=j}} \begin{pmatrix} 1 & \xi_i & \xi_i^2 \\ \xi_i & \xi_i^2 & \xi_i^3 \\ \xi_i^2 & \xi_i^3 & \xi_i^4 \end{pmatrix} \right) > 0 \quad \text{for } j = 1, 2. \end{aligned}$$

Then the “ignore- \hat{F} ” estimator of the parameters of two lines is strongly consistent, that is,

$$\hat{k}_j \rightarrow k_j, \quad \hat{h}_j \rightarrow h_j, \quad j = 1, 2,$$

as $n \rightarrow \infty$ almost surely.

Proposition 6. *If in the structural model, $k_1 < k_2$, $E|\xi_1|^3 < \infty$, and neither condition (13) nor condition (14) holds, then the “ignore- \hat{F} ” estimator is consistent.*

Now, we state the asymptotic normality of the “ignore- \hat{F} ” estimator.

Proposition 7. *In the functional model, assume the following:*

- $k_1 < k_2$,
- for $j = 1, 2$ and $p = 0, 1, \dots, 6$, the following limits exist and are finite:

$$\mu_p^{(j)} := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\substack{i=1, \dots, n \\ \nu(i)=j}} \xi_i^p.$$

- for $j = 1$ and $j = 2$, the matrices

$$\begin{pmatrix} \mu_0^{(j)} & \mu_1^{(j)} & \mu_2^{(j)} \\ \mu_1^{(j)} & \mu_2^{(j)} & \mu_3^{(j)} \\ \mu_2^{(j)} & \mu_3^{(j)} & \mu_4^{(j)} \end{pmatrix}$$

are nonsingular.

Then the “ignore- \hat{F} ” estimator $(\hat{k}_1, \hat{h}_1, \hat{k}_2, \hat{h}_2)^\top$ is asymptotically normal, namely

$$\sqrt{n} \begin{pmatrix} \hat{k}_1 - k_1 \\ \hat{h}_1 - h_1 \\ \hat{k}_2 - k_2 \\ \hat{h}_2 - h_2 \end{pmatrix} \xrightarrow{d} N(0, K \Sigma_{\tilde{\beta}} K^\top), \quad (21)$$

where $\Sigma_{\tilde{\beta}}$ is the asymptotic covariance matrix of $\tilde{\beta}$, and K is the 4×6 matrix of derivatives of the mapping $(A, 2B, C, 2D, 2E, F)^\top \mapsto (k_1, h_1, k_2, h_2)^\top$ defined in (15)–(17) at the true parameters β_{tn} , that is,

$$K = \left. \frac{d \log(\beta)}{d\beta^\top} \right|_{\beta=\beta_{\text{tn}}}.$$

The matrix $K \Sigma_{\tilde{\beta}} K^\top$ is nonsingular.

Proposition 8. *If, in the structural model, $k_1 < k_2$, $E \xi_1^6 < \infty$, and neither (13) nor (14) holds, then the “ignore- \hat{F} ” estimator is asymptotically normal, that is, (21) holds.*

Remark 2. The estimators \hat{k}_1 , \hat{h}_1 , \hat{k}_2 , and \hat{h}_2 obtained in (18)–(20) do not change if \hat{A} , \hat{B} , \dots , \hat{E} are multiplied by a common factor. So it does not matter which normalization of β is used.

Equation (11) represents a couple of intersecting straight lines if and only if

$$\begin{vmatrix} A & B & D \\ B & C & E \\ D & E & F \end{vmatrix} = 0 \quad \text{and} \quad AC < B^2. \quad (22)$$

Denote

$$\Delta(\beta) = \begin{vmatrix} A & B & D \\ B & C & E \\ D & E & F \end{vmatrix} = ACF + 2BDE - AE^2 - CD^2 - B^2F,$$

$$\Delta'(\beta) = \frac{d\Delta(\beta)}{d\beta^\top} = (CF - E^2, DE - BF, AF - D^2, BE - CD, BD - AE, AC - B^2),$$

where the function $\Delta(\beta)$ and its derivative $\Delta'(\beta)$ are evaluated at the point $\beta = (A, 2B, C, 2D, 2E, F)^\top$.

Perform one-step update of the estimator $\tilde{\beta}$ to make it closer to the surface $\Delta(\beta) = 0$:

$$\tilde{\beta}_{1\text{st}} = \tilde{\beta} - \frac{\Delta(\tilde{\beta})}{\Delta'(\tilde{\beta}) \hat{\Sigma}_{\tilde{\beta}} \Delta'(\tilde{\beta})^\top} \hat{\Sigma}_{\tilde{\beta}} \Delta'(\tilde{\beta})^\top. \quad (23)$$

Then use expressions (15)–(17) to estimate k_1, h_1, k_2, h_2 :

$$(\hat{k}_{1,1\text{st}}, \hat{h}_{1,1\text{st}}, \hat{k}_{2,1\text{st}}, \hat{h}_{2,1\text{st}})^\top = \log(\tilde{\beta}_{1\text{st}}).$$

Proposition 9. *Under the conditions of Proposition 7 in the functional model or under the conditions of Proposition 8 in the structural model, the estimator $(\hat{k}_{1,1st}, \hat{h}_{1,1st}, \hat{k}_{2,1st}, \hat{h}_{2,1st})^\top$ is consistent and asymptotically normal, and its asymptotic covariance matrix is equal to*

$$K \left(\Sigma_{\hat{\beta}} - \frac{\Sigma_{\hat{\beta}} \Delta'(\beta_{tn})^\top \Delta'(\beta_{tn}) \Sigma_{\hat{\beta}}}{\Delta'(\beta_{tn}) \Sigma_{\hat{\beta}} \Delta'(\beta_{tn})^\top} \right) K^\top.$$

Remark 3. The normalization of the estimator $\hat{\beta}$ affects its asymptotic covariance matrix, and hence has effect on the estimates $(\hat{k}_{1,1st}, \hat{h}_{1,1st}, \hat{k}_{2,1st}, \hat{h}_{2,1st})^\top$. However, the normalization does not affect the asymptotic covariance matrix of $(\hat{k}_{1,1st}, \hat{h}_{1,1st}, \hat{k}_{2,1st}, \hat{h}_{2,1st})^\top$.

2.2 Orthogonal regression estimator

The sum of squared distances between each observed point and the closer of two lines is equal to

$$Q(k_1, h_1, k_2, h_2) = \sum_{i=1}^n \min \left(\frac{(y_i - k_1 x_i - h_1)^2}{k_1^2 + 1}, \frac{(y_i - k_2 x_i - h_2)^2}{k_2^2 + 1} \right). \quad (24)$$

The *orthogonal regression estimator* is a Borel-measurable function of observations such that

$$(\hat{k}_{1,OR}, \hat{h}_{1,OR}, \hat{k}_{2,OR}, \hat{h}_{2,OR}) \in \underset{(k_1, h_1, k_2, h_2) \in \mathbb{R}^4}{\operatorname{argmax}} Q(k_1, h_1, k_2, h_2).$$

In the functional model, the orthogonal regression estimator is the maximum likelihood estimator. However, because the dimension of parameter space grows as the sample size is increasing, the orthogonal regression estimator may be inconsistent.

2.3 Parametric maximum likelihood estimator

The estimator is constructed in the structural model, so it should be called the structural maximum likelihood estimator.

If a Gaussian distribution of a random point (ξ, η) is concentrated on a straight line $\eta = k\xi + h$, then it is a singular normal distribution:

$$(\xi, \eta) \sim N \left(\begin{pmatrix} \mu_\xi \\ k\mu_\xi + h \end{pmatrix}, \begin{pmatrix} \sigma_\xi^2 & k\sigma_\xi^2 \\ k\sigma_\xi^2 & k^2\sigma_\xi^2 \end{pmatrix} \right), \quad (25)$$

where μ_ξ and σ_ξ^2 are the expectation and variance of the random variable ξ . Note that the covariance matrix $\sigma_\xi^2 \begin{pmatrix} 1 & k \\ k & k^2 \end{pmatrix}$ is singular and positive semidefinite.

If the distribution of a random point (ξ_i, η_i) is concentrated on two straight lines $\eta = k_1\xi + h_1$ and $\eta = k_2\xi + h_2$ and the distribution on each line is Gaussian, then, due to (25), the conditional distributions are

$$[(\xi_i, \eta_i) \mid \nu(i) = j] \sim N \left(\begin{pmatrix} \mu_{j\xi} \\ k_j\mu_{j\xi} + h_j \end{pmatrix}, \begin{pmatrix} \sigma_{j\xi}^2 & k_j\sigma_{j\xi}^2 \\ k_j\sigma_{j\xi}^2 & k_j^2\sigma_{j\xi}^2 \end{pmatrix} \right) = N(\mu_j, \Sigma_{0j})$$

for $j = 1, 2$. The matrices Σ_{0j} are positive semidefinite and singular, that is, $\lambda_{\min}(\Sigma_{01}) = \lambda_{\min}(\Sigma_{02}) = 0$, and the points μ_j are the centers of Gaussian distribution of the points on each line.

The distribution of (ξ_i, η_i) is a mixture of two singular normal distributions

$$\begin{pmatrix} \xi_i \\ \eta_i \end{pmatrix} \sim \text{mixture of } \begin{cases} N(\mu_1, \Sigma_{01}) & \text{with weight } p, \\ N(\mu_2, \Sigma_{02}) & \text{with weight } 1 - p, \end{cases} \quad (26)$$

where $p = P(\nu(i) = 1) = P(\nu(1) = 1)$ is the probability that the point (ξ_i, η_i) lies on the first line.

The distribution of the observed points is also a mixture of two Gaussian distributions

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim \text{mixture of } \begin{cases} N(\mu_1, \Sigma_1) & \text{with weight } p, \\ N(\mu_2, \Sigma_2) & \text{with weight } 1 - p, \end{cases} \quad (27)$$

with $\Sigma_j = \Sigma_{0j} + \sigma^2 \mathbf{I}$, where σ^2 is the error variance; see (4). Note that $\lambda_{\min}(\Sigma_1) = \lambda_{\min}(\Sigma_2) = \sigma^2$.

The likelihood function for the sample of points with a mixture of two normal distributions is

$$L(p, \mu_1, \Sigma_1, \mu_2, \Sigma_2) = \prod_{i=1}^n (p \phi_{N(\mu_1, \Sigma_1)}(x_i, y_i) + (1-p) \phi_{N(\mu_2, \Sigma_2)}(x_i, y_i)), \quad (28)$$

where

$$\phi_{N(\mu, \Sigma)}(x, y) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right)^\top \Sigma^{-1} \left(\begin{pmatrix} x \\ y \end{pmatrix} - \mu \right) \right\}$$

is the density of a bivariate normal distribution.

One method of evaluating the maximum likelihood estimator is as follows:

1. Find the point of conditional minimum

$$(\hat{\mu}_1, \hat{\Sigma}_1, \hat{\mu}_2, \hat{\Sigma}_2) = \underset{\substack{\mu_1, \Sigma_1, \mu_2, \Sigma_2 \\ \text{such that } \lambda_{\min}(\Sigma_1) = \lambda_{\min}(\Sigma_2)}}{\operatorname{argmin}} \min_{p \in [0, 1]} L(p, \mu_1, \Sigma_1, \mu_2, \Sigma_2).$$

2. Set

$$\hat{\sigma}^2 = \lambda_{\min}(\Sigma_1) = \frac{1}{2} (\hat{\sigma}_{1xx} + \hat{\sigma}_{1yy} - \sqrt{(\hat{\sigma}_{1xx} - \hat{\sigma}_{1yy})^2 + 4\hat{\sigma}_{1xy}^2}). \quad (29)$$

Here $\hat{\sigma}_{jxx}$, $\hat{\sigma}_{jxy}$, and $\hat{\sigma}_{jyy}$ are the entries of the matrix $\hat{\Sigma}_j$, and $\hat{\mu}_{jx}$ and $\hat{\mu}_{jy}$ are the elements of the vector $\hat{\mu}_j$:

$$\hat{\mu}_j = \begin{pmatrix} \hat{\mu}_{jx} \\ \hat{\mu}_{jy} \end{pmatrix}, \quad \hat{\Sigma}_j = \begin{pmatrix} \hat{\sigma}_{jxx} & \hat{\sigma}_{jxy} \\ \hat{\sigma}_{jxy} & \hat{\sigma}_{jyy} \end{pmatrix}.$$

3. Find the estimates $\hat{k}_1, \hat{h}_1, \hat{k}_2, \hat{h}_2$ from the equations

$$\hat{\boldsymbol{\mu}}_j = \begin{pmatrix} \mu_{jx} \\ \hat{k}_j \mu_{jx} + \hat{h}_j \end{pmatrix}, \quad \hat{\boldsymbol{\Sigma}}_j = \begin{pmatrix} \hat{\sigma}_{j\xi}^2 + \hat{\sigma}^2 & \hat{k}_j \hat{\sigma}_{j\xi}^2 \\ \hat{k}_j \hat{\sigma}_{j\xi}^2 & \hat{k}_j^2 \hat{\sigma}_{j\xi}^2 + \hat{\sigma}^2 \end{pmatrix},$$

that is, set

$$\hat{k}_j = \frac{\hat{\sigma}_{jxy}}{\hat{\sigma}_{jxx} - \hat{\sigma}^2}, \quad (30)$$

$$\hat{h}_j = \hat{\mu}_{jy} - \hat{k}_j \hat{\mu}_{jx}. \quad (31)$$

The denominator $\hat{\sigma}_{jxx} - \hat{\sigma}^2$ may be equal to 0 with some positive probability. Occurrence of this event means that the estimated figure is a straight line and a single point outside the line rather than two straight lines.

In order to make the statement of consistency easier, assume that $k_1 < k_2$ and choose the estimator such that $\hat{k}_1 \leq \hat{k}_2$.

2.4 RBAN moment estimator

The *regular best asymptotically normal* (RBAN) estimators were developed by Chiang [4]. Our RBAN moment estimator differs from the original RBAN so that not only the observed points (x_i, y_i) , but also monomials $x_i^p y_i^q$, $p + q \leq 4$, are averaged.

Introduce the 14-dimensional vectors whose elements are the monomials of coordinates of observed points:

$$m(x, y) = (x^4, x^3y, x^2y^2, xy^3, y^4, x^3, x^2y, xy^2, y^3, x^2, xy, y^2, x, y)^\top, \quad (32)$$

$$\mathbf{m}_i = m(x_i, y_i).$$

Evaluate the average and sample covariance matrix of the vectors \mathbf{m}_i :

$$\bar{\mathbf{m}} = \frac{1}{n} \sum_{i=1}^n \mathbf{m}_i, \quad \boldsymbol{\Sigma}_m = \frac{1}{n} \sum_{i=1}^n (\mathbf{m}_i - \bar{\mathbf{m}})(\mathbf{m}_i - \bar{\mathbf{m}})^\top.$$

Denote

$$f_1(k, h, \sigma^2; (\mu_p)_{q=1}^4) = \mathbb{E} m(\xi + \delta, k\xi + h + \varepsilon),$$

where ξ, δ , and ε are independent random variables such that

$$\mathbb{E} \xi^q = \mu_q \quad \text{and} \quad (\delta, \varepsilon)^\top \sim N(0, \sigma^2 \mathbf{I}).$$

Basically, the function f_1 is defined for all μ_p , $p = 1, \dots, 4$, that comprise possible 4-tuples of moments of a random variable, that is, satisfy

$$\mu_2 - \mu_1^2 \geq 0,$$

$$(\mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4)(\mu_2 - \mu_1^2) - (\mu_3 - 3\mu_2\mu_1 + 2\mu_1^3)^2 - (\mu_2 - \mu_1^2)^3 \geq 0;$$

see [10]. However, since the elements of the vector-function f_1 are polynomials of its arguments, it can be extended to \mathbb{R}^7 .

Denote

$$\begin{aligned} f_2(k_1, h_1, k_2, h_2, \sigma^2; p, (\mu_q^{(j)})_{j=1, q=1}^2 \quad 4) \\ = pf_1(k_1, h_1, \sigma^2; (\mu_q^{(1)})_{q=1}^4) + (1-p)f_1(k_2, h_2, \sigma^2; (\mu_q^{(2)})_{q=1}^4). \end{aligned}$$

In the structural model,

$$\mathbb{E} \mathbf{m}_i = f_2(k_1, h_1, k_2, h_2, \sigma^2; P(\nu(1)=1), (\mathbb{E}[\xi_1^q \mid \nu(1)=j])_{j=1, q=1}^2 \quad 4).$$

Consider the equation

$$f_2(\hat{k}_1, \hat{h}_1, \hat{k}_2, \hat{h}_2, \hat{\sigma}^2; \hat{p}, (\hat{\mu}_q^{(j)})_{j=1, q=1}^2 \quad 4) = \overline{\mathbf{m}}. \quad (33)$$

It is a system of 14 equations in 14 variables. If (33) has a solution, then the moment estimator can be defined as one of the solutions. However, (33) may have no solution.

In the rest of Section 2.4, $\mu_{\bullet}^{(\bullet)} = (\mu_q^{(j)})_{j=1, q=1}^2 \quad 4$ is a 2×4 matrix.

The estimator is defined as a point where $(f_2(\dots) - \overline{\mathbf{m}})^\top \Sigma_{\overline{\mathbf{m}}}^{-1} (f_2(\dots) - \overline{\mathbf{m}})$ attains its minimum:

$$\begin{aligned} & (\hat{k}_1, \hat{h}_1, \hat{k}_2, \hat{h}_2, \hat{\sigma}^2) \\ & = \operatorname{argmin}_{k_1, \dots, \sigma^2} \min_{p, \mu_{\bullet}^{(\bullet)}} (f_2(k_1, \dots, \sigma^2; p, \mu_{\bullet}^{(\bullet)}) - \overline{\mathbf{m}})^\top \Sigma_{\overline{\mathbf{m}}}^{-1} (f_2(k_1, \dots, \sigma^2; p, \mu_{\bullet}^{(\bullet)}) - \overline{\mathbf{m}}). \end{aligned}$$

This minimization problem is similar to that in Theorem 6 in [4]. The minimum

$$\min_{p \in \mathbb{R}} \min_{\mu_{\bullet}^{(\bullet)} \in \mathbb{R}^{2 \times 4}} (f_2(k_1, \dots, \sigma^2; p, \mu_{\bullet}^{(\bullet)}) - \overline{\mathbf{m}})^\top \Sigma_{\overline{\mathbf{m}}}^{-1} (f_2(k_1, \dots, \sigma^2; p, \mu_{\bullet}^{(\bullet)}) - \overline{\mathbf{m}})^\top$$

can be evaluated explicitly, and this allows us to reduce the dimension of minimization problem. The reduction of dimension of the optimization problem was used, for example, in [8].

The routines evaluating the RBAN-moment estimator and the estimator for its covariance matrix are developed without rigid theoretical basis; see Section 4.3.

3 Equivariance

3.1 Two definitions of equivariance

The similarity transformation of \mathbb{R}^2 is

$$g(\mathbf{z}) = K\mathbf{U}\mathbf{z} + \Delta\mathbf{z}, \quad \mathbf{z} \in \mathbb{R}^2, \quad (34)$$

where \mathbf{U} is an orthogonal matrix, $K \neq 0$ is a scaling coefficient, and $\Delta\mathbf{z} \in \mathbb{R}^2$ is an intercept.

The transformation of a sample of points acting elementwise is also denoted $g(Z)$: if $Z = \{\mathbf{z}_i, i = 1, \dots, n\}$, then $g(Z) = \{g(\mathbf{z}_i), i = 1, \dots, n\}$.

Hereafter, we use vector notation: the observed points are denoted $\mathbf{z}_i = (x_i, y_i)^\top$, and the true points are denoted $\boldsymbol{\zeta}_i = (\xi_i, \eta_i)^\top$.

The underlying statistical structure is $(\mathbb{R}^{n \times 2}, \mathcal{B}(\mathbb{R}^{n \times 2}), P_{Z|\theta}, \theta \in \Theta)$, where $Z \in \mathbb{R}^{n \times 2}$ is the observed sample, $Z = \{z_i, i = 1, \dots, n\}$, $\mathcal{B}(\mathbb{R}^{n \times 2})$ is the Borel σ -field, and θ is a parameter that uniquely identifies the distribution of the observed points; $\theta = (\zeta_1, \dots, \zeta_n; \sigma^2)$ in the functional model, and $\theta = (P_\zeta; \sigma^2)$ in the structural model. Here ζ_1, \dots, ζ_n are points located on two strait lines, and P_ζ is a probability measure concentrated on two straight lines.

The statistical structure is invariant with respect to transformation g if the change of the probability measure induced by the transformation of the sample can be obtained by some transformation \tilde{g} of parameters, that is, if there exists a bijection $\tilde{g} : \Theta \rightarrow \Theta$ such that

$$\forall \theta \in \Theta : P_{g(Z)|\theta} = P_{Z|\tilde{g}(\theta)}.$$

Here $P_{g(Z)|\theta}$ is the induced probability measure; it is sometimes denoted $P_{g(Z)|\theta} = P_{Z|\theta g^{-1}}$.

The statistical structure is similarity invariant if it is invariant with respect to all similarity transformations of the form (34).

In order to become similarity invariant, the underlying statistical structure needs some extension. We assume that the true points lie on two lines, which *may* be parallel to the y -axis. The following restrictions do not ruin the invariance:

- The true lines ℓ_1 and ℓ_2 intersect each other but do not coincide.
- The true points ζ_1, \dots, ζ_n in the functional model or the set $\text{supp}(P_\zeta)$ where the true points are concentrated in the structural model can be covered by two lines uniquely. In the structural model, this means that the nonidentifiability conditions (13) and (14) do not hold.

With these restrictions, the statistical structure is invariant with

$$\tilde{g}(\zeta_1, \dots, \zeta_n; \sigma^2) = (g(\zeta_1), \dots, g(\zeta_n); K^2 \sigma^2)$$

in the functional model and

$$\tilde{g}(P_\zeta; \sigma^2) = (P_{g(\zeta)}; K^2 \sigma^2)$$

in the structural model.

Let $\ell\ell(\theta)$ and $\text{sigma2}(\theta)$ be functions that extract the parameters of interest. If, in the functional model, $\theta = (\zeta_1, \dots, \zeta_n; \sigma^2)$ and points ζ_1, \dots, ζ_n lie on the lines ℓ_1 and ℓ_2 or if, in the structural model, $\theta = (P_\zeta; \sigma^2)$ and the probability measure is concentrated on the union of two lines $\ell_1 \cup \ell_2$, then $\ell\ell(\theta) = \{\ell_1, \ell_2\}$. If $\theta = (\dots; \sigma^2)$, then $\text{sigma2}(\theta) = \sigma^2$.

We treat $\{\ell_1, \ell_2\}$ as an unordered couple, that is, $\{\ell_1, \ell_2\} = \{\ell_2, \ell_1\}$.

The transformation of the lines parameters and the transformation of σ^2 do not interfere each other, and these transformations are not interfered by a particular location or distribution of true points on the lines, that is, the parameter transformation \tilde{g} is such that there exists transformations $\tilde{g}_{\ell\ell}$ and \tilde{g}_{σ^2} such that

$$\begin{aligned} \ell\ell(\tilde{g}(\theta)) &= \tilde{g}_{\ell\ell}(\ell\ell(\theta)), \\ \text{sigma2}(\tilde{g}(\theta)) &= \tilde{g}_{\sigma^2}(\text{sigma2}(\theta)). \end{aligned}$$

These transformations are

$$\begin{aligned}\tilde{g}_{\ell\ell}(\{\ell_1, \ell_2\}) &= \{g(\ell_1), g(\ell_2)\}, \\ \tilde{g}_{\sigma^2}(\sigma^2) &= K^2\sigma^2.\end{aligned}\tag{35}$$

The estimator is called equivariant with respect to the transformation g if, when the data are transformed, the estimator follows the inducing transformation of parameters. The estimator $\widehat{\ell\ell}(Z)$ for two lines and the estimator $\hat{\sigma}^2(Z)$ for error variance are equivariant with respect to similarity transformation g if

$$\widehat{\ell\ell}(g(Z)) = \tilde{g}_{\ell\ell}(\widehat{\ell\ell}(Z)),\tag{36}$$

$$\hat{\sigma}^2(g(Z)) = \tilde{g}_{\sigma^2}(\hat{\sigma}^2(Z)).\tag{37}$$

The estimator is called similarity equivariant if it is equivariant with respect to any similarity transformation g .

In a fitting problem, an estimator for a “true figure” is called *fitting equivariant* with respect to transformation $g(z)$, $g : \mathbb{R} \rightarrow \mathbb{R}$ if, when the sample is transformed, the estimated “true figure” follows the same transformation g . An estimator is called *similarity fitting equivariant* if it is *fitting equivariant* with respect to any similarity transformation.

In the two-line fitting problem, denote by $\cup\{\ell_1, \ell_2\} = \ell_1 \cup \ell_2$ the union of a pair of two lines. An estimator $\widehat{\ell\ell}(Z)$ is similarity fitting equivariant if and only if for any similarity transformation $g(z)$,

$$\cup \widehat{\ell\ell}(g(Z)) = g(\cup \widehat{\ell\ell}(Z)).\tag{38}$$

The similarity fitting equivariant estimator depends on geometry of the plane and does not depend on the Cartesian coordinate system used.

Because of (35), in the two-line fitting model, the estimator for two lines $\widehat{\ell\ell}(Z)$ is similarity equivariant if and only if it is similarity fitting equivariant.

3.2 Similarity equivariance of the five estimators

Some troubles, which may arise during estimation, are not addressed yet.

- The estimation may fail with small positive probability. For example, the conic section estimated with the ALS2 estimator is an ellipse with some positive probability, and if it is, then the “ignore- \widehat{F} ” estimator fails. (If the estimator is consistent, then the failure probability tends to 0 as $n \rightarrow \infty$).
- The estimation may fail, for example, because the estimated line should be parallel to the y -axis, but the estimating procedure does not handle such case.
- The optimization problem may have multiple extremal points. For the ALS2 estimator, it may occur that $\dim\{\beta : \Psi_n(\hat{\sigma}^2)\beta = 0\} > 1$.

In order to define the equivariance of an unreliable estimator, we allow that the estimators fail simultaneously in both sides of (36), (37), or (38). Also, we allow that

for fixed similarity transformation $g(z)$, equation (36), (37), or (38) does not hold with probability 0.

The equivariance of the ALS2 estimator in the conic section fitting problem is verified in [11, Section 5.5] (see Theorem 30 there for similarity fitting equivariance). That implies the equivariance of the “ignore- \hat{F} ” estimator.

In order to make the *updated before ignore- \hat{F} step* estimator equivariant, we use normalization of the ALS2 estimator (9) rather than $\|\hat{\beta}\| = 1$.

The *orthogonal regression* estimator and the parametric *maximum likelihood* estimator are maximum likelihood estimators, but in different models. Thus, they are equivariant.

The criterion function for the *RBAN-moment* estimator is similarity invariant. This means that the criterion function does not change when the data sample follows a similarity transformation and the parameters follow the inducing transformation. Thus, the RBAN-moment estimator is equivariant.

3.3 An example of equivariant but not fitting equivariant estimator

Consider a further restriction of the mixture-of-two-normal-distributions model from Section 2.3. Assume that covariance matrices of Σ_1 and Σ_2 have the same diagonal entries but additive inverse off-diagonal entries:

$$\Sigma_1 = \begin{pmatrix} \sigma_\xi^2 + \sigma^2 & -k\sigma_\xi^2 \\ -k\sigma_\xi^2 & k^2\sigma_\xi^2 + \sigma^2 \end{pmatrix} \quad \text{and} \quad \Sigma_2 = \begin{pmatrix} \sigma_\xi^2 + \sigma^2 & k\sigma_\xi^2 \\ k\sigma_\xi^2 & k^2\sigma_\xi^2 + \sigma^2 \end{pmatrix}.$$

The statistical structure is invariant in scaling of the y -coordinate, $(x_{\text{new}}, y_{\text{new}}) = (x_{\text{old}}, ry_{\text{old}})$, $r > 0$. This transformation maps the lines $y = -kx + h_1$ and $y = kx + h_2$ onto the lines $y = -rkx + rh_1$ and $y = rkx + rh_2$, respectively. The maximum likelihood estimator in this model is equivariant. However, this equivariance is somewhat strange. The transformation of parameters that induces the scaling of the y -coordinate of the observed points does not induce the same transformation of the true points nor the same mapping of the true lines. The estimated lines follow the transformation of parameters rather than the transformation of observed points. This is illustrated in Fig. 1.

Let k_{old} be the true value of the parameter k before the transformation. Then after the transformation, the value of the parameter is

$$k_{\text{new}} = \frac{t + \sqrt{t^2 + 4r^2 k_{\text{old}}^2 \sigma_{\xi \text{ old}}^4}}{2rk_{\text{old}} \sigma_{\xi \text{ old}}^2}$$

with $t = (r^2 k_{\text{old}}^2 - 1) \sigma_{\xi \text{ old}}^2 + (r^2 - 1) \sigma_{\text{old}}^2$. If $0 < r^2 \neq 1$, $k_{\text{old}} \neq 0$, and $\sigma^2 > 0$, then $k_{\text{new}} \neq rk_{\text{old}}$. Hence, the maximum likelihood estimator is not fitting equivariant with respect to scaling of the y -coordinate here.

4 Simulations

4.1 Simulation setup

A sample of the true points (ξ_i, η_i) , $i = 1, \dots, n$, is generated from a random distribution concentrated on (a subset of) two lines. Three distributions of the true points are used; see Fig. 2:

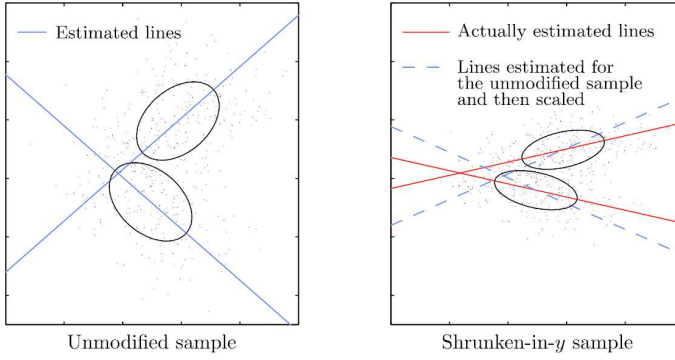


Fig. 1. Two samples, one of points (x, y_{old}) (the unmodified sample) and one of points $(x, y_{\text{new}}) = (x, \frac{1}{2}y_{\text{old}})$ (the shrunk-in- y sample), are fitted with two lines (the estimated lines are the solid lines on the figures). The estimated lines for the unmodified sample (blue solid lines on the left figure) when scaled with the same transformation as the observed points are scaled (blue dashed line on the right figure) do not coincide with the actually estimated lines for the shrunk-in- y sample (red solid lines on the right figure). The ellipsoids are the 40% ellipsoids of the estimated normal distributions (the compound distributions of the estimated mixture distributions) (color figure online)

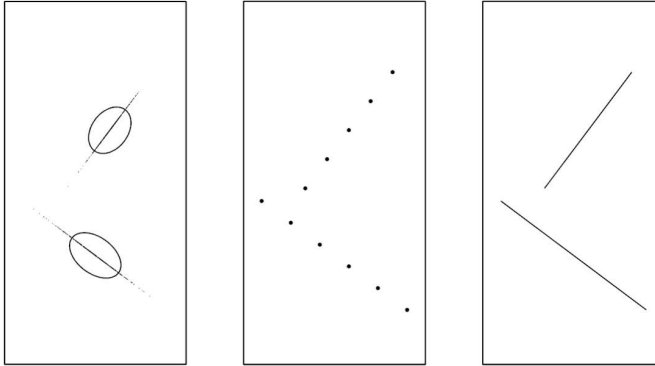


Fig. 2. Three distributions of the true points: a mixture of two singular normal distributions, a discrete distribution, and a uniform distribution on two line segments. For the first case, a sample of 1000 points is plotted, whereas for the second and third cases, the support of the distribution of the true points is plotted. For the first case, the distribution of the *observed* points is a mixture of normal distributions, and 40% ellipsoids for its components are plotted

- a mixture of two singular normal distributions,
- a discrete distribution,
- a uniform distribution on two line segments.

These three distributions of true points are concentrated on the same two lines

$$4y = 1 - 3x \quad \text{and} \quad 12y = 16x + 5,$$

which intersect one another at the point $(-0.08, 0.31)$.

For the same sample of true points $\{(\xi_i, \eta_i), i = 1, \dots, n\}$, 100 samples of the measurement errors $\{(\delta_i, \varepsilon_i), i = 1, \dots, n\}$, $(\delta_i, \varepsilon_i)^\top \sim N(0, \sigma^2)$, are simulated, and 100 samples of the observed points (x_i, y_i) are obtained; see (2) and (3). For each sample of the observed points, the estimates of the parameters of the true lines were evaluated with the following five methods: two ALS2-based estimators (the ignore- \hat{F} estimator and the estimator with one-step update of the ALS2 estimator before the ignore- \hat{F} step), the orthogonal regression estimator, the parametric maximum likelihood estimator, and the RBAN moment estimator.

For each estimated couple of lines, the point of their intersection is found. The 100 estimates of intersection points are averaged, and their sample standard deviations are evaluated. For the ALS2-based estimators and the RBAN moment estimator, the standard errors of the estimators are also evaluated.

4.2 Notes on computation of particular estimators

For computation of the *orthogonal regression* estimator, the *k*-means method is used. Initially, two lines were chosen randomly. Then *classification* and *mean* steps are alternated. On the classification step, the observed points are split into two clusters based on which line is closer to the point. (The first cluster contains all the observed points that are closer to the first line than to the second line, and the second cluster contains the other observed points.) On the means step, each cluster is fitted with a straight line by the orthogonal regression method (the two lines are updated). The algorithm is completed when the classification step does now change the clusters. The obtained parameters of the two lines deliver a local minimum to the criterion function $Q(k_1, h_1, k_2, h_2)$ (24). Trying to obtain the global minimum, the algorithm is restarted several times with different initial two lines.

For computation of the *parametric maximum likelihood estimator*, the expectation-maximization algorithm [2] is used. The equation for optimization problem of finding a minimum of the likelihood function $L(p, \mu_1, \Sigma_1, \mu_2, \Sigma_2)$ (28) such that $\lambda_{\min}(\Sigma_1) = \lambda_{\min}(\Sigma_2)$ is

$$\begin{aligned} \sum_{i=1}^n \frac{\phi_{N(\mu_1, \Sigma_1)}(x_i, y_i) - \phi_{N(\mu_2, \Sigma_2)}(x_i, y_i)}{p\phi_{N(\mu_1, \Sigma_1)}(x_i, y_i) + (1-p)\phi_{N(\mu_2, \Sigma_2)}(x_i, y_i)} &= 0, \\ \sum_{i=1}^n \frac{p \frac{\partial \phi_{N(\mu_1, \Sigma_1)}(x_i, y_i)}{\partial \text{par}} + (1-p) \frac{\partial \phi_{N(\mu_2, \Sigma_2)}(x_i, y_i)}{\partial \text{par}}}{p\phi_{N(\mu_1, \Sigma_1)}(x_i, y_i) + (1-p)\phi_{N(\mu_2, \Sigma_2)}(x_i, y_i)} &= 0, \end{aligned}$$

where *par* is a vector parameterization of $(\mu_1, \Sigma_1, \mu_2, \Sigma_2)$ such that $\lambda_{\min}(\Sigma_1) = \lambda_{\min}(\Sigma_2)$. Hence, the maximum likelihood estimator is a stationary point of the function

$$Q_{\mathbf{w}}(p, \text{par}) = \sum_{i=1}^n (w_i \ln(p\phi_{N(\mu_1, \Sigma_1)}(x_i, y_i)) + (1-w_i) \ln((1-p)\phi_{N(\mu_1, \Sigma_1)}(x_i, y_i)))$$

with fixed

$$w_i = \frac{\hat{p}\phi_{N(\hat{\mu}_1, \hat{\Sigma}_1)}(x_i, y_i)}{\hat{p}\phi_{N(\hat{\mu}_1, \hat{\Sigma}_1)}(x_i, y_i) + (1-\hat{p})\phi_{N(\hat{\mu}_2, \hat{\Sigma}_2)}(x_i, y_i)}.$$

The EM algorithm is iterative. Once the $(m-1)$ th approximation $(p^{(m-1)}, \mu_1^{(m-1)}, \Sigma_1^{(m-1)}, \mu_2^{(m-1)}, \Sigma_2^{(m-1)})$ is obtained, the weights are evaluated:

$$w_i^{(m-1)} = \frac{p^{(m-1)} \phi_{N(\mu_1^{(m-1)}, \Sigma_1^{(m-1)})}(x_i, y_i)}{p^{(m-1)} \phi_{N(\mu_1^{(m-1)}, \Sigma_1^{(m-1)})}(x_i, y_i) + (1-p^{(m-1)}) \phi_{N(\mu_2^{(m-1)}, \Sigma_2^{(m-1)})}(x_i, y_i)}.$$

Then m th approximation $(p^{(m)}, \mu_1^{(m)}, \Sigma_1^{(m)}, \mu_2^{(m)}, \Sigma_2^{(m)})$ is obtained by minimizing

$$\sum_{i=1}^n (w_i^{(m-1)} \ln(p \phi_{N(\mu_1, \Sigma_1)}(x_i, y_i)) + (1-w_i^{(m-1)}) \ln((1-p) \phi_{N(\mu_2, \Sigma_2)}(x_i, y_i)))$$

under the constraint $\lambda_{\min}(\Sigma_1) = \lambda_{\min}(\Sigma_2)$. The point where the minimum is attained can be explicitly expressed in $w_i^{(m-1)}$, x_i , and y_i , $i = 1, \dots, n$.

4.3 RBAN-moment estimator

In case the criterion function

$$\begin{aligned} Q(\theta) &= Q(k_1, h_1, k_2, h_2, \sigma^2) \\ &= \min_{p \in \mathbb{R}} \min_{\mathbf{M} \in \mathbb{R}^{2 \times 4}} (f_2(\theta; p, \mathbf{M}) - \overline{\mathbf{m}})^\top \Sigma_m^{-1} (f_2(\theta; p, \mathbf{M}) - \overline{\mathbf{m}}) \end{aligned}$$

has multiple minima, a consistent estimator—that is, the “ignore- \widehat{F} ” estimator—is used as the initial point, and the criterion function $Q(\theta)$ is searched for a local minimum nearby. Here $\theta = (k_1, h_1, k_2, h_2, \sigma^2)^\top$ is a vector meaning the parameters of interest.

The knowledge or misspecification of the parameter p does not affect the estimator for the parameters of interest k_1, \dots, σ^2 . Thus, for estimation of the asymptotic covariance matrix, assume $p = 0.5$ to be known. The estimator of the asymptotic covariance matrix of (θ, \mathbf{M}) is

$$\widehat{\Sigma}_{\theta, \mathbf{M}} = (f'_2(\hat{\theta}; 0.5, \widehat{\mathbf{M}})^\top \Sigma_m^{-1} f'_2(\hat{\theta}; 0.5, \widehat{\mathbf{M}}))^{-1},$$

where

$$f'_2(\theta; 0.5, \mathbf{M}) = \left(\frac{\partial f_2(\theta; 0.5, \mathbf{M})}{\partial \theta^\top}, \frac{\partial f_2(\theta; 0.5, \mathbf{M})}{\partial (\text{vec } \mathbf{M})^\top} \right).$$

The estimator of the asymptotic covariance matrix of θ is the principal submatrix of $\widehat{\Sigma}_{\theta, \mathbf{M}}$.

4.4 Simulation results

Average of estimated centers over 100 simulations, standard deviations over 100 simulations, and medians of estimated standard errors are presented in Tables 1–3.

Using the estimator \widehat{F} (by *one-step update* before ignore- \widehat{F} step), we improve the precision of estimation. The precision of the *RBAN-moment* estimator approximates the precision of the *updated before ignore- \widehat{F} step* estimator.

Table 1. Means, standard deviations, and median standard errors of the estimates of intersection points for true points having mixture of singular normal distributions

Method	Means		Standard deviations		Standard errors	
True value	-0.08	0.31				
$n = 1000, \quad \sigma = 0.1 \quad (\sigma^2 = 0.01)$						
Ignore- \hat{F}	-0.1098	0.2753	0.8611	0.3866	0.1918	0.1125
Update	-0.0820	0.2912	0.0706	0.0620	0.0437	0.0479
OR	0.6533	3.4524	0.0877	0.6783		
ML	-0.0795	0.3077	0.0326	0.0269		
RBAN	0.0647	0.3759	0.3563	0.2606	0.0350	0.0438
$n = 10000, \quad \sigma = 0.1 \quad (\sigma^2 = 0.01)$						
Ignore- \hat{F}	-0.0909	0.3052	0.0646	0.0308	0.0601	0.0303
Update	-0.0796	0.3080	0.0127	0.0156	0.0124	0.0155
OR	0.5492	3.1488	0.0175	0.1701		
ML	-0.0776	0.3083	0.0103	0.0088		
RBAN	-0.0789	0.3100	0.0126	0.0154	0.0127	0.0154
$n = 100000, \quad \sigma = 0.1$						
Ignore- \hat{F}	-0.0799	0.3101	0.0211	0.0095	0.0188	0.0093
Update	-0.0801	0.3098	0.0037	0.0042	0.0039	0.0047
OR	0.5606	3.2041	0.0063	0.0484		
ML	-0.0801	0.3101	0.0030	0.0025		
RBAN	-0.0801	0.3101	0.0038	0.0042	0.0039	0.0048
$n = 1000, \quad \sigma = 0.02$						
Ignore- \hat{F}	-0.0799	0.3099	0.0151	0.0075	0.0147	0.0072
Update	-0.0795	0.3097	0.0052	0.0051	0.0052	0.0049
OR	-0.0792	0.3092	0.0050	0.0044		
ML	-0.0794	0.3093	0.0048	0.0043		
RBAN	-0.0797	0.3098	0.0063	0.0057	0.0052	0.0049

The parametric *maximum likelihood* estimator is the best when the normality condition, which was assumed during construction of the estimator, is satisfied. Otherwise, it is biased.

The *orthogonal regression* and the *maximum likelihood* estimators are good for small error variance ($\sigma^2 = 0.02^2$). For $\sigma^2 = 0.1^2$, the *orthogonal regression* estimator is broken down when the distribution of true points is a mixture of two normal distributions and is biased for the two other distributions of true points.

Mean-square deviance of the intersection of the estimated lines from the true intersection point is presented in Table 4.

For small errors, the *RBAN-moment* estimator is a bit less accurate than the *updated before ignore- \hat{F}* step estimator. For $\sigma^2 = 0.1^2$, the difference is negligible.

The parametric *maximum likelihood* estimator has the smallest deviation from the true value, except for the discrete distribution of true points and $\sigma^2 = 0.01$.

For small errors ($\sigma^2 = 0.02^2$), the *orthogonal regression* estimator outperforms the consistent estimators and has the deviation approximately as small as the parametric *maximum likelihood* estimator.

Normalization of the estimator of β affects the ALS2-based estimator of two lines with *one-step update before the ignore- \hat{F}* step. With normalization $\|\hat{\beta}\| = 1$, the derived estimator of two lines is not equivariant, whereas with normalization

Table 2. Means, standard deviations, and median standard errors of the estimates of intersection points for discrete distribution of the true points

Method	Means		Standard deviations		Standard errors	
True value	-0.08	0.31				
<hr/>						
$n = 1000, \quad \sigma = 0.1$						
Ignore- \hat{F}	-0.0699	0.3077	0.0263	0.0290	0.0241	0.0263
Update	-0.0722	0.3116	0.0197	0.0186	0.0203	0.0175
OR	-0.0755	0.3188	0.0148	0.0144		
ML	-0.0958	0.3315	0.0131	0.0120		
RBAN	-0.0717	0.3105	0.0209	0.0175	0.0205	0.0178
<hr/>						
$n = 10000, \quad \sigma = 0.1$						
Ignore- \hat{F}	-0.0783	0.3109	0.0092	0.0078	0.0080	0.0083
Update	-0.0785	0.3114	0.0071	0.0054	0.0065	0.0061
OR	-0.0721	0.3157	0.0048	0.0046		
ML	-0.0931	0.3278	0.0043	0.0035		
RBAN	-0.0786	0.3113	0.0071	0.0053	0.0065	0.0061
<hr/>						
$n = 100000, \quad \sigma = 0.1$						
Ignore- \hat{F}	-0.0798	0.3098	0.0031	0.0024	0.0026	0.0027
Update	-0.0799	0.3099	0.0025	0.0016	0.0021	0.0019
OR	-0.0715	0.3151	0.0017	0.0013		
ML	-0.0932	0.3283	0.0013	0.0011		
RBAN	-0.0799	0.3099	0.0024	0.0017	0.0021	0.0019
<hr/>						
$n = 1000, \quad \sigma = 0.02$						
Ignore- \hat{F}	-0.0796	0.3094	0.0033	0.0032	0.0036	0.0033
Update	-0.0798	0.3097	0.0030	0.0024	0.0033	0.0023
OR	-0.0782	0.3086	0.0021	0.0018		
ML	-0.0786	0.3087	0.0019	0.0018		
RBAN	-0.0796	0.3092	0.0030	0.0028	0.0033	0.0024

$\tilde{\beta}^\top \Psi'_n(\hat{\sigma}^2) \tilde{\beta} = -n$, the derived estimator is equivariant. Comparison of equivariant and nonequivariant versions of the estimator is displayed in Table 5.

There is a tendency that the equivariant version of the estimator is more accurate for small samples than the nonequivariant version. The two versions of the estimator are consistent and asymptotically equivalent. When the estimation is precise, the difference between the versions is negligible. When the estimation is imprecise, it is impossible to make inference which version is more accurate.

4.5 Comparison of two estimators for asymptotic covariance matrix in the conic section fitting model

In [13] a conic section fitting model is considered, and two estimators ($\hat{\Sigma}_{\text{true}}$ and $\hat{\Sigma}_{\text{sample}}$) for the asymptotic covariance matrix of the ALS2 estimator are constructed.

The software developed here can be used to make numerical comparison of the estimates of the asymptotic covariance matrices. The data are generated as described in Section 4.1 with 1000 simulations for each set of true points. Thus, the true conic unnecessarily was chosen degenerate. For each simulation, the parameters of the conic section were estimated; its center is found, and two confidence ellipsoids for the center were constructed using two different estimators of the asymptotic covariance matrix.

Table 3. Means, standard deviations, and median standard errors of the estimates of intersection points for uniform distribution of the true points on two line segments

Method	Means		Standard deviations		Standard errors	
True value	-0.08	0.31				
<i>n</i> = 1000, $\sigma = 0.1$						
Ignore- \hat{F}	-0.0785	0.3122	0.0363	0.0274	0.0318	0.0301
Update	-0.0794	0.3140	0.0216	0.0258	0.0205	0.0290
OR	-0.0616	0.3127	0.0185	0.0167		
ML	-0.0934	0.3118	0.0116	0.0111		
RBAN	-0.0807	0.3126	0.0219	0.0293	0.0193	0.0292
<i>n</i> = 10000, $\sigma = 0.1$						
Ignore- \hat{F}	-0.0796	0.3107	0.0103	0.0103	0.0099	0.0095
Update	-0.0796	0.3110	0.0067	0.0103	0.0065	0.0094
OR	-0.0639	0.3087	0.0064	0.0049		
ML	-0.0904	0.3106	0.0042	0.0033		
RBAN	-0.0797	0.3107	0.0066	0.0104	0.0064	0.0095
<i>n</i> = 100000, $\sigma = 0.1$						
Ignore- \hat{F}	-0.0798	0.3098	0.0035	0.0030	0.0032	0.0030
Update	-0.0798	0.3098	0.0021	0.0029	0.0020	0.0030
OR	-0.0625	0.3085	0.0015	0.0014		
ML	-0.0891	0.3107	0.0012	0.0011		
RBAN	-0.0796	0.3097	0.0023	0.0030	0.0020	0.0030
<i>n</i> = 1000, $\sigma = 0.02$						
Ignore- \hat{F}	-0.0799	0.3100	0.0041	0.0032	0.0041	0.0035
Update	-0.0798	0.3101	0.0033	0.0032	0.0032	0.0034
OR	-0.0803	0.3103	0.0023	0.0021		
ML	-0.0805	0.3101	0.0022	0.0020		
RBAN	-0.0798	0.3100	0.0035	0.0033	0.0032	0.0034

Table 4. Mean-square distances between estimated and true intersection points

<i>n</i>	σ	Ignore- \hat{F}	Update	OR	ML	RBAN
<i>Distribution of true points is a mixture of normals</i>						
1000	0.1	0.9403	0.0954	3.2978	0.0421	0.6124
10000	0.1	0.0722	0.0201	2.9127	0.0138	0.0199
100000	0.1	0.0230	0.0056	2.9645	0.0038	0.0056
1000	0.02	0.0168	0.0073	0.0067	0.0065	0.0084
<i>Discrete distribution of true points</i>						
1000	0.1	0.0403	0.0281	0.0228	0.0320	0.0284
10000	0.1	0.0121	0.0091	0.0118	0.0228	0.0090
100000	0.1	0.0039	0.0029	0.0101	0.0226	0.0029
1000	0.02	0.0046	0.0038	0.0036	0.0032	0.0042
<i>Uniform distribution of true points</i>						
1000	0.1	0.0453	0.0367	0.0310	0.0209	0.0365
10000	0.1	0.0145	0.0123	0.0181	0.0117	0.0123
100000	0.1	0.0046	0.0036	0.0177	0.0093	0.0037
1000	0.02	0.0052	0.0046	0.0031	0.0030	0.0048

The sample coverage probability and median (over 1000 ellipsoids) area of the confidence ellipsoids is presented in Table 6. The ellipsoids were constructed for confidence levels 0.8 and 0.95. The area of 95% confidence ellipsoids is displayed in

Table 5. Comparison of two versions (equivariant (ev) and nonequivariant (ne)) of the updated before ignore- \hat{F} step estimator

n	σ	Ver.	Means		Standard deviations		Standard errors	
True value:			-0.08	0.31				
<i>Distribution of true points is a mixture of normals</i>								
1000	0.1	ev	-0.082046	0.291175	0.070617	0.062003	0.043713	0.047939
		ne	-0.044038	0.247382	0.251372	0.514473	0.038853	0.050167
10000	0.1	ev	-0.079623	0.308039	0.012652	0.015582	0.012403	0.015471
		ne	-0.085055	0.304177	0.015924	0.018695	0.012506	0.015550
100000	0.1	ev	-0.080137	0.309780	0.003710	0.004173	0.003925	0.004749
		ne	-0.080991	0.309386	0.003880	0.004255	0.003926	0.004742
1000	0.02	ev	-0.079548	0.309703	0.005156	0.005131	0.005174	0.004891
		ne	-0.079918	0.309508	0.005247	0.005149	0.005179	0.004918
<i>Discrete distribution of true points</i>								
1000	0.1	ev	-0.072202	0.311648	0.019709	0.018553	0.020266	0.017500
		ne	-0.071460	0.312230	0.020049	0.018740	0.020213	0.017457
10000	0.1	ev	-0.078482	0.311377	0.007087	0.005371	0.006518	0.006066
		ne	-0.078418	0.311436	0.007090	0.005387	0.006520	0.006054
100000	0.1	ev	-0.079868	0.309929	0.002460	0.001647	0.002060	0.001900
		ne	-0.079863	0.309934	0.002461	0.001647	0.002060	0.001901
1000	0.02	ev	-0.079772	0.309728	0.002967	0.002376	0.003320	0.002344
		ne	-0.079755	0.309732	0.002963	0.002375	0.003319	0.002339
<i>Uniform distribution of true points</i>								
1000	0.1	ev	-0.079405	0.313977	0.021551	0.025759	0.020451	0.028992
		ne	-0.078507	0.315350	0.022219	0.026091	0.020512	0.029115
10000	0.1	ev	-0.079604	0.311024	0.006673	0.010337	0.006467	0.009389
		ne	-0.079576	0.311176	0.006685	0.010349	0.006456	0.009372
100000	0.1	ev	-0.079795	0.309802	0.002075	0.002919	0.001974	0.002994
		ne	-0.079794	0.309818	0.002076	0.002921	0.001972	0.002992
1000	0.02	ev	-0.079833	0.310081	0.003252	0.003249	0.003172	0.003418
		ne	-0.079825	0.310097	0.003250	0.003249	0.003169	0.003411

Table 6. Coverage probability and area of confidence ellipsoids (c.e.) for centers by the ALS2 estimator

n	σ	$\hat{\Sigma}_{\text{true}}$ -based estimator			$\hat{\Sigma}_{\text{sample}}$ -based estimator		
		Coverage probab.		Area of 95% c.e., $\times 10^{-4}$	Coverage probab.		Area of 95% c.e., $\times 10^{-4}$
		80%,	95%,		80%,	95%,	
		%	%		%	%	
<i>Distribution of true points is a mixture of normals</i>							
1000	0.1	70.6	80.2	1449.	70.0	79.2	1562.
10000	0.1	79.4	93.8	236.2	79.6	92.9	259.4
100000	0.1	80.7	94.9	15.38	80.6	94.9	15.17
1000	0.02	80.4	95.1	15.95	78.0	94.1	19.86
<i>Discrete distribution of true points</i>							
1000	0.1	78.1	93.9	81.80	77.4	93.4	78.94
10000	0.1	81.1	95.6	12.34	80.9	95.8	12.39
100000	0.1	80.1	94.6	1.205	79.9	94.7	1.204
1000	0.02	81.0	94.9	1.984	81.3	95.2	1.988
<i>Uniform distribution of true points</i>							
1000	0.1	82.1	94.3	152.9	81.5	94.3	138.4
10000	0.1	81.0	96.6	20.36	80.3	96.3	18.92
100000	0.1	78.4	95.0	1.823	78.6	95.0	1.842
1000	0.02	78.7	94.7	2.926	78.5	94.1	3.041

Table 6, and the area of 80% confidence ellipsoid is $\log_{20}(5) = 0.5372$ of the area of 95% confidence ellipsoids.

Note that standard errors for coverage probability are 1.3% for 80% confidence ellipsoids and 0.7% for 95% confidence ellipsoids. The simulations do not allow us to make an inference which estimator is better. Thus, $\hat{\Sigma}_{\text{sample}}$ -based estimator *updated before ignore- \hat{F} step* is compared with other estimators in simulations in Section 4.4 because of simpler explicit expression for $\hat{\Sigma}_{\text{sample}}$.

A Proofs

Proof of Proposition 1. The strong consistency of the estimator follows from [11, Theorem 17]. Under the conditions of Proposition 1,

$$\frac{1}{n} \Psi_n(\sigma^2) \rightarrow \bar{\Psi}_\infty \quad \text{a.s.} \quad (39)$$

$$\frac{1}{n} \bar{\Psi}_n \rightarrow \bar{\Psi}_\infty \quad (\text{a.s. in the structural model}). \quad (40)$$

By Lemma 5 in [11], $\limsup \frac{1}{n} \beta^\top \bar{\Psi}'_n \beta < 0$, which, together with (40), implies $\beta^\top \bar{\Psi}'_\infty \beta < 0$ (see the proof of Theorem 2 in [12]). Then

$$\frac{\hat{\beta}^\top \Psi'_n(\hat{\sigma}^2) \hat{\beta}}{n \|\hat{\beta}\|^2} = \frac{\hat{\beta}^\top \Psi'_n(\sigma^2) \hat{\beta}}{n \|\hat{\beta}\|^2} + \frac{(\hat{\sigma}^2 - \sigma^2) \hat{\beta}^\top \Psi'_1 \hat{\beta}}{\|\hat{\beta}\|^2} \rightarrow \frac{\beta^\top \bar{\Psi}'_\infty \beta}{\|\beta\|^2} \quad \text{a.s.} \quad (41)$$

Eventually, the left-hand side of (41) is negative. \square

Proof of Proposition 2. The strong consistency of $\tilde{\beta}$ follows from (7) and (41). The proof of asymptotic normality and consistency of the estimator of the asymptotic covariance matrix can be obtained by modification of the proofs of Theorem 2 in [12] and Theorem 3 in [13]. \square

Proof of Proposition 3. The conditions of consistency Theorem 1 in [12] can be verified, and the consistency follows.

The most tedious is the condition

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \lambda_{\min, 2}(\bar{\Psi}_n) > 0. \quad (42)$$

Denote

$$K_j = \begin{pmatrix} 0 & 0 & 1 \\ 0 & h_j & k_j \\ 0 & 1 & 0 \\ h_j & k_j & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad j = 1, 2.$$

Then $K_1 K_1^\top + K_2 K_2^\top$ is a positive semidefinite matrix, and

$$\det(K_1 K_1^\top + K_2 K_2^\top) = 2(h_1 - h_2)^4 + 2(h_1 - h_2)^2(k_1 - k_2)^2 + 2(k_1 - k_2)^4 > 0.$$

Thus, $\lambda_{\min}(K_1 K_1^\top + K_2 K_2^\top) > 0$.

The matrix

$$K_1 \sum_{\substack{i=1, \dots, n \\ \nu(i)=1}} \begin{pmatrix} 1 & \xi_i & \xi_i^2 \\ \xi_i & \xi_i^2 & \xi_i^3 \\ \xi_i^2 & \xi_i^3 & \xi_i^4 \end{pmatrix} K_1^\top + K_2 \sum_{\substack{i=1, \dots, n \\ \nu(i)=2}} \begin{pmatrix} 1 & \xi_i & \xi_i^2 \\ \xi_i & \xi_i^2 & \xi_i^3 \\ \xi_i^2 & \xi_i^3 & \xi_i^4 \end{pmatrix} K_2^\top$$

is the principal submatrix of $\bar{\Psi}_n$. By the Cauchy interlacing theorem,

$$\begin{aligned} \lambda_{\min, 2}(\bar{\Psi}_n) &\geq \lambda_{\min} \left(\sum_{j=1}^2 K_j \sum_{\substack{i=1, \dots, n \\ \nu(i)=j}} \begin{pmatrix} 1 & \xi_i & \xi_i^2 \\ \xi_i & \xi_i^2 & \xi_i^3 \\ \xi_i^2 & \xi_i^3 & \xi_i^4 \end{pmatrix} K_j^\top \right) \\ &\geq \lambda_{\min}(K_1 K_1^\top + K_2 K_2^\top) \min_{j=1, 2} \left(\lambda_{\min} \left(\sum_{\substack{i=1, \dots, n \\ \nu(i)=j}} \begin{pmatrix} 1 & \xi_i & \xi_i^2 \\ \xi_i & \xi_i^2 & \xi_i^3 \\ \xi_i^2 & \xi_i^3 & \xi_i^4 \end{pmatrix} \right) \right), \end{aligned}$$

and then inequality (42) can easily be proved. \square

Proof of Proposition 4. Proposition 4 follows from Proposition 25 in [11]. The identifiability condition (SS-) in [11] holds because the intersection of a couple of lines and a conic section may be a finite set with not more than four points, a straight line, a straight line and a point outside the line, or the couple of lines and the conic section coincide; in the last case, the coefficients of the equations for the lines and the conic section satisfy relations (12). \square

Proofs of Propositions 5, 6, 7, and 8. Consistence of the “ignore- \hat{F} ” estimator follows from the consistency of the ALS2 estimator $\hat{\beta}$ and from the continuity of the function $\text{lob}(\beta)$ at the point of the true value of the parameter β . The asymptotic normality of the “ignore- \hat{F} ” follows from the asymptotic normality of $\tilde{\beta}$ and the differentiability of $\text{lob}(\beta)$ at the point β_{tn} . \square

Proof of Proposition 9. The consistency and asymptotic normality of the $\tilde{\beta}$ estimator, the differentiability of the functional $\Delta(\beta)$ at point β_{tn} (note that $\Delta(\beta_{\text{tn}}) = 0$), and the convergence

$$\frac{1}{\Delta'(\tilde{\beta}) \hat{\Sigma}_{\tilde{\beta}} \Delta'(\tilde{\beta})^\top} \hat{\Sigma}_{\tilde{\beta}} \Delta'(\tilde{\beta})^\top \rightarrow \frac{1}{\Delta'(\beta_{\text{tn}}) \Sigma_{\tilde{\beta}} \Delta'(\beta_{\text{tn}})^\top} \Sigma_{\tilde{\beta}} \Delta'(\beta_{\text{tn}})^\top$$

imply the convergence and asymptotic normality of the updated estimator $\tilde{\beta}_{1\text{st}}$. Thus, the consistency and asymptotic normality of $(\hat{k}_{1,1\text{st}}, \hat{h}_{1,1\text{st}}, \hat{k}_{2,1\text{st}}, \hat{h}_{2,1\text{st}})^\top$ can be proved similarly to those of the “ignore- \hat{F} ” estimator. \square

References

- [1] Ahn, S.J.: Least Squares Orthogonal Distance Fitting of Curves and Surfaces in Space. Springer, Heidelberg (2004). doi:[10.1007/b104017](https://doi.org/10.1007/b104017)

- [2] Bilmes, J.A.: A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Technical report TR-97-021, International Computer Science Institute (1998)
- [3] Cheng, C.-L., Van Ness, J.W.: Statistical Regression with Measurement Error. Arnold, London (1999). [MR1719513](#)
- [4] Chiang, C.L.: On regular best asymptotically normal estimates. *Ann. Math. Stat.* **27**(2), 336–351 (1956). doi:[10.1214/aoms/1177728262](#) [MR0089558](#)
- [5] Fazekas, I., Kukush, A., Zwanzig, S.: Correction of nonlinear orthogonal regression estimator. *Ukr. Math. J.* **56**(8), 1308–1330 (2004). [MR2136312](#). doi:[10.1007/s11253-005-0059-0](#)
- [6] Kukush, A., Markovsky, I., Van Huffel, S.: Consistent fundamental matrix estimation in a quadratic measurement error model arising in motion analysis. *Comput. Stat. Data Anal.* **41**(1), 3–18 (2002). [MR1944689](#). doi:[10.1016/S0167-9473\(02\)00068-3](#)
- [7] Kukush, A., Markovsky, I., Van Huffel, S.: Correction of nonlinear orthogonal regression estimator. *Comput. Stat. Data Anal.* **47**(1), 123–147 (2004). [MR2087933](#). doi:[10.1016/j.csda.2003.10.022](#)
- [8] Markovsky, I., Van Huffel, S., Kukush, A.: On the computation of the multivariate structured total least squares estimator. *Numer. Linear Algebra Appl.* **11**(5–6), 591–608 (2004). [MR2067822](#). doi:[10.1002/nla.361](#)
- [9] Repetatska, G.: An improved orthogonal regression estimator for the implicit functional errors-in-variables model. *Bull. Kyiv Natl. Taras Shevchenko Univ. Math. Mech.* **23**, 37–45 (2010)
- [10] Rohatgi, V.K., Székely, G.J.: Sharp inequalities between skewness and kurtosis. *Stat. Probab. Lett.* **8**(4), 296–299 (1989). [MR1028986](#). doi:[10.1016/0167-7152\(89\)90035-7](#)
- [11] Shklyar, S., Kukush, A., Markovsky, I., Van Huffel, S.: On the conic section fitting problem. *J. Multivar. Anal.* **98**(3), 588–624 (2007). [MR2293016](#). doi:[10.1016/j.jmva.2005.12.003](#)
- [12] Shklyar, S.V.: Singular asymptotic normality of an estimator in the conic section fitting problem. I. *Teor. Imovir. Mat. Stat.* **92**, 137–150 (2015)
- [13] Shklyar, S.V.: Singular asymptotic normality of an estimator in the conic section fitting problem. II. *Teor. Imovir. Mat. Stat.* **93**, 163–180 (2015)
- [14] Vidal, R., Ma, Y., Sastry, S.: Generalized principal component analysis (GPCA). *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(12), 1945–1959 (2005). doi:[10.1109/TPAMI.2005.244](#)
- [15] Waibel, P., Matthes, J., Gröll, L.: Constrained ellipse fitting with center on a line. *J. Math. Imaging Vis.* **53**(3), 364–382 (2015). [MR3397105](#). doi:[10.1007/s10851-015-0584-x](#)
- [16] Zelnik-Manor, L., Irani, M.: Multi-view subspace constraints on homographies. In: *Proceedings of the 7th IEEE International Conference on Computer Vision*, 1999, vol. 2, pp. 710–7152 (1999). doi:[10.1109/ICCV.1999.790291](#)